

Moteurs de recherche: mode d'emploi

Retour 

1. Trouver les bons mots clés

Vous vous en êtes certainement rendu compte : l'une des plus dures "épreuves", lorsque l'on recherche une information sur le Web, consiste à trouver les bons mots clés, ceux qui vont caractériser au mieux sa recherche. La tâche n'est pas si compliquée : il suffit pour être efficace de se plier à une certaine stratégie.

Prendre son temps

Pour la plupart des internautes en quête d'une info en particulier, le premier réflexe consiste à lancer un outil de recherche (moteur ou annuaire) et à taper les quelques mots qui viennent immédiatement à l'esprit. C'est une erreur. Avant toute considération syntaxique, la précipitation constitue le premier écueil à éviter. Les moteurs de recherche, pour puissants qu'ils sont, n'ont pas la capacité de synthèse et d''analyse d'un cerveau humain. Il nous incombe donc de leur pré-mâcher le travail en nous livrant à une petite gymnastique mentale du choix des mots clés.

Essayez-donc de ne pas vous précipiter sur votre clavier. Prenez le temps qu'il faut, cinq ou dix minutes si nécessaire, avant de vous lancer dans un surf effréné. Réfléchissez bien à ce que vous recherchez, posez-vous les questions essentielles sur les objectifs à atteindre, bref, remuez-vous les méninges, Google ne le fera pas pour vous. Cela peut paraître évident, mais ces quelques minutes de réflexion prises au début de votre quête vous feront gagner beaucoup de temps par la suite.

Discerner les trois grandes familles de mots

Essayez de bien définir les mots clés que vous allez utiliser. N'hésitez pas à les noter sur une feuille (oui, de papier, vous vous souvenez..?). Globalement, les mots clés peuvent être classés en trois grandes familles :

a) Mots clés "primaires" ou "larges":

Ils définissent, de façon globale, le champ d'investigation. Exemple : agro-alimentaire, finance, commerce, santé, cinéma, loisirs, voyages, tourisme, littérature, histoire, sciences, etc.

b) Mots clés "secondaires" ou "profonds":

Au sein du domaine global décrit précédemment, ils définissent la requête elle-même.

Exemple : jus d'orange, marché, bourse, Chicago (recherche : le cours du jus d'orange à la bourse de Chicago), marignan, 1515, François 1er, Léon X (pour des renseignements sur la bataille de Marignan), Montréal, Québec, Canada, hôtels, hébergement (pour trouver où dormir à Montréal), etc.

c) Mots clés "interdits":

Il s'agit de mots qui peuvent vous induire en erreur lors de vos recherches et donc à exclure, comme indiqué ici.

Exemple : vous recherchez des informations sur le marché de la pomme aux Etats-Unis. Mot clé primaire : food. Mots clés secondaires : apple, study, market. Mot clé interdit : computer. Car les ordinateurs de la marque Apple ne vous intéressent pas, en tout cas pas dans cette recherche. Autre exemple : la même recherche dans le monde francophone. Mot clé primaire : agro-alimentaire. Mots clés secondaires : pomme, marché, étude, France (par exemple). Mot clé interdit : Adam. Car la pomme d'Adam ne fait pas non plus l'objet de vos investigations.

Une fois notés au moins quatre mots clés (mais n'hésitez pas à en inscrire une bonne dizaine sur

votre bout de papier), vous voilà mûr pour consulter votre outil de recherche habituel. Les requêtes efficaces devraient maintenant être plus faciles et plus rapides à identifier.

2. Utilisation du signe "+" pour forcer la recherche d'un mot

Quel que soit l'objet de votre recherche, on peut affirmer que plusieurs mots clé valent mieux qu'un. C'est vrai dans l'absolu, mais encore faut-il respecter la syntaxe adéquate lorsque l'on multiplie les pistes d'investigation.

L'erreur habituelle

Pour lancer une recherche sur plusieurs mots, le réflexe commun consiste à les saisir à la file, en les séparant d'un simple espace comme on le ferait en langage naturel.

exemple : pour chercher des sites relatifs à l'Olympique de Marseille (le club de foot phocéen), on a tendance à saisir "football marseille".

C'est ce qu'il ne faut surtout pas faire. En effet, pour l'immense majorité des outils de recherche, **l'espace équivaut non pas à un ET mais à un OU**. Ainsi la requête de notre exemple sera interprété de cette façon :

"Je demande les pages qui contiennent le terme football ou les pages qui contiennent le terme marseille ou celles qui contiennent les deux" !

En somme, en commettant cette petite erreur, on élargit involontairement la recherche au lieu de l'affiner.

Ce qu'il faut faire

Il sufft de faire précéder les mots recherchés du signe "+" pour les lier par un ET. Ainsi dans notre exemple, pour s'assurer de la présence conjointe des mots marseille et football dans les réponses proposés par les moteurs, il faudra rédiger la requête comme suit :

exemple : "+football +marseille".

La demande porte alors uniquement sur les pages qui contiennent les deux mots en excluant toutes celles qui ne comporte que l'un des deux.

La syntaxe

Cette syntaxe sera comprise de la plupart des moteurs de recherche actuels, mais il importe de respecter trois points :

- **le signe + doit être "collé"** au mot qui le suit, sans espace entre les deux.
- **Un espace** doit tout de même séparer les deux mots.
- **Chaque mot exigé doit être accompagné d'un "+"**.

exemple : si l'on écrit "football +marseille", le mot marseille sera obligatoirement présent dans les réponses, mas pas le terme football puisqu'aucun signe "+" ne lui est associé.

3. Exclure des mots à l'aide du signe "-"

De même qu'il est possible d'imposer la recherche d'un mot en lui adjoignant le signe mathématique "+", on peut demander aux moteur de recherche d'éliminer systématiquement toutes les réponses qui comportent un ou plusieurs termes. Pour ce faire, on aura recours au signe "-" (moins), autrement dit on fera précéder le mot d'un simple tiret.

Exemple : vous recherchez des articles sur le cancer.

L'horoscope des gens nés sous le signe du cancer ne vous intéresse pas, or sur le Web les pages traitant d'astrologie sont au moins aussi nombreuses que les documents médicaux... Il faut donc restreindre le champ de la requête, comme suit :
"cancer -horoscope"

Précisons que ce signe "-" doit impérativement être collé au mot à exclure. La plupart des moteurs de recherche actuels sauront ainsi l'interpréter comme l'opérateur logique SAUF. En d'autres termes, l'ordre ainsi exprimé signifie pour eux : "***recherche les pages ou les sites contenant le mot "cancer" et ne contenant pas le mot "horoscope"***".

Requêtes complexes

Les syntaxes peuvent bien entendu être mixées avec les guillemets (voir expression exacte) et l'opérateur ET (le signe "+"), notre requête précédente pouvant tout à fait s'énoncer sous sous la forme

+cancer +santé -horoscope +"rhume des foins" -horoscope

Exercices

Voici deux petits exercices pour vérifier vos acquis.
Quelle est la différence entre ces deux requêtes :

cancer -horoscope	**+cancer -horoscope**

Réponse : il n'y a aucune différence. Dans les deux cas, "cancer" est le terme expressément recherché tandis que le mot clé "horoscope" est interdit. Dans cet exemple, le signe "+" est du reste inutile.
Le même exercice, un poil plus relevé :

cancer sein -horoscope	**+cancer +sein -horoscope**

Réponse : il y a cette fois une différence notable. Les mots "cancer" et "sein" sont liés dans le premier cas par un OU (on recherchera alors cancer OU sein OU les deux), dans le deuxième cas par un ET (seule la présence simultanée des deux termes sera détectée).

La manipulation de ces opérateurs logiques est un jeu d'enfant. Quoi qu'il en soit, plus vous utiliserez d'opérateurs comme ET (+) ou SAUF (-) et moins les moteurs de recherche vous retourneront des réponses hors de propos.

4. Les guillemets pour chercher une expression exacte

Lors d'une recherche sur le Web, il arrive souvent que l'on ait à mener ses investigations sur une expression comprenant plusieurs mots séparés par un espace. Par exemple : New York, Euro 2000, Festival d'Avignon, etc. **Pour retrouver une expression exacte, il suffit de la saisir entre guillemets**. Dès lors votre moteur de recherche habituel ne vous retournera que les occurences rigoureusement identiques de cette suite de mot. En d'autres termes, les requêtes "New York", "Euro 2000" ou "Festival d'Avignon", sont interprétés comme l'ordre de rechercher " les pages ou les sites qui contiennent les mots demandés, tels que présentés, l'un à côté de l'autre et dans cet ordre".

Explications

La saisie de New York (les deux mots séparés par un espace) recherchera soit le mot New

seulement, soit le mot York seulement, soit les deux termes dans la base de données appropriée (annuaire ou moteur).

La saisie de +New +York induira la recherche simultanée des deux mots, mais l'un pouvant se trouver au début du document et l'autre à la fin. Donc, la présence des deux mots est nécessaire (opérateur ET), mais pas obligatoirement l'un à côté de l'autre.

Par contre, **la saisie de "New York"** entre guillemets provoquera une recherche avec les deux mots demandés, mais cette fois obligatoirement l'un à côté de l'autre. Il s'agit donc de la syntaxe la plus restrictive des trois.

Précautions

Les guillemets sont fortement recommandés pour rechercher de façon plus fine une expression, mais la demande effectuée auprès du moteur devient alors très stricte et sans dérivés possibles. Il ne faut donc pas en abuser, la moindre petite faute de frappe, et le moteur rentrerait bredouille.

Par exemple, la saisie de "festival d'Avignon" ne trouvera pas l'expression "festival à Avignon". Et "coupe du monde" ne trouvera pas "coupes du monde", avec un pluriel sur le premier mot.

5. La question des caractères accentués

La question n'a aucune importance pour les anglophones dont la langue ne comporte pas d'accent, mais pour nous autres qui parlons français (avec une cédille...) elle est cruciale. Alors, **faut-il oui ou non saisir les mots clés avec des lettres accentuées ?**

Chaque moteur de recherche gère les caractères spéciaux (accents, cédille...) à sa façon, nous vous renvoyons à la documentation de votre outil de recherche fétiche pour connaître son comportement aide pour déterminer , mais on peut dégager trois cas de figure :

- soit **ils ne tiennent pas compte de l'accentuation** : la saisie du mot "bébé" est équivalente à celle de "bebe" ou "bebé"

- soit, **ils trouvent toutes les occurrences si le mot demandé est non accentuée**. Exemple : "bebe" trouve "bebe", "bébé" ou "bebé". Mais, dès qu'une lettre accentuée est présente dans le terme de la requête, c'est ce terme-là, et pas un autre, qui sera trouvé : "bébé" trouve alors "bébé" mais plus "bebe" ou "bebé".

- soit, c'est **toujours la graphie exacte** qui est recherchée : "bebe" trouve "bebe" et rien d'autre, "bébé" trouve "bébé" et rien d'autre, etc.

Nos conseils dans l'absolu

La situation est donc assez complexe et totalement dépendante des outils utilisés. Aussi, **si vous êtes sûr de l'orthographe exacte du mot recherché, tapez-le en lettres accentuées**. Au pire, vous éliminerez les pages qui le contiennent avec une faute d'orthographe.
En revanche, **si vous avez un doute sur la façon dont s'écrit le mot à rechercher, saisissez le plutôt sans accents**. Si vous tombez sur un moteur de la troisième catégorie évoquée (recherche de la graphie exacte), vous vous en apercevrez immédiatement puisqu'il ne vous retournera aucun résultat. Il sera alors temps d'ouvrir un dictionnaire...

6. Comment les moteurs de recherche classent-ils leurs résultats ?

Mieux utiliser les outils de recherche, c'est également mieux comprendre comment ils fonctionnent. Et, notamment, comment ils classent les informations fournies lorsque vous recherche une information en leur sein. Après les annuaires la semaine dernière, on passe aux moteurs de recherche...

Nous avons vu, la semaine dernière, comment les annuaires classaient les résultats qu'ils proposaient aux internautes.

En ce qui concerne les moteur de recherche, qui effectuent leurs investigations dans plusieurs centaines de millions de pages Web, la situation est plus complexe. En effet, de nombreux paramètres sont pris en compte pour effectuer le calcul de pertinence. Citons :

* la présence du mot demandé dans le titre du document;

* sa présence dans le texte de la page, mais cela ne suffit pas : une page contenant le mot demandé en haut du document sera mieux "notée" qu'une autre le présentant en bas de page. De nombreux critères sur la façon dont le texte est affiché sont ainsi pris en compte, comme la proximité, la mise en exergue (gras, taille des caractères, etc.);

* la présence du mot demandé dans certains champs qui n'apparaissent pas sur la page mais qui sont fournis aux outils de recherche. On appelle ces champs les "balises Meta" ;

* sa présence dans l'adresse de la page (exemple : www.bidule.fr/dossier/france2.html)

* l'indice de popularité du document (le nombre de liens "pointant" vers cette page sur le Web).

Les critères de pertinence sont très nombreux pour les moteurs de recherche. Historiquement, ils se sont basés, pour classer leurs documents, sur des critères dits "in the page", c'est-à-dire contenus dans le document à classer : titre, texte visible, balises Meta, etc.

Nouveaux critères : "off the page"

Mais ces critères n'ont rapidement pas suffi, car trop de "petis malins" ont essayé de les détourner à leur avantage pour voir leur site bien classé sur ces outils. Aussi, les moteurs de recherche ont imaginé de nouveaux critères, dits "off the page". Ceux-ci sont alors pris en compte, non plus dans le contenu de la page, mais dans son environnement. L'indice de popularité a été le premier d'entre eux. Autre exemple de critère "off the page" : l'indice de clic, symbolisé par le système DirectHit (http://www.directhit.com/).

Pour bien comprendre comment il fonctionne, analysons le comportement "classique" d'un internaute devant un moteur de recherche. Il va sur la page d'accueil, saisit un ou plusieurs mots dans un formulaire, consulte la page de résultats proposée (sur laquelle plusieurs liens sont indiqués, classés par ordre de pertinence), il choisit l'un d'entre eux, va sur le site correspondant, le consulte. Si cette page ne lui convient pas, il revient sur la page de résultats du moteur (par le bouton "précédent" du navigateur), choisit un autre lien parmi ceux proposés, jusqu'à ce qu'il ait trouvé son bonheur.

DirectHit va, en fait, fonctionner en tâche de fond sur un moteur existant. À chaque consultation d'un internaute, il va noter sur quel lien il a cliqué et quel était le rang (le classement) de ce lien. Il calcule ensuite combien de temps l'utilisateur met avant de revenir sur la page de résultats. S'il ne revient pas, il en "déduit" que le site proposé était a priori pertinent. Son adresse sera alors mieux classée dans les résultats suivants, lors d'une interrogation sur le même mot-clé. Et ainsi de suite, les interrogations et la façon d'interroger et de naviguer des internautes vont alors enrichir la base de données de DirectHit. Ingénieux, non ?

On peut le voir, le système de classement des moteurs de recherche est basé sur des calculs de pertinence associés à de très nombreux critères. Et tout cela en quelques dizièmes de secondes. Plutôt pas mal pour trier parmi plusieurs centaines de milions de pages Web, non ?

Récapitulatif pour le moteur « Google »



Présentation générale

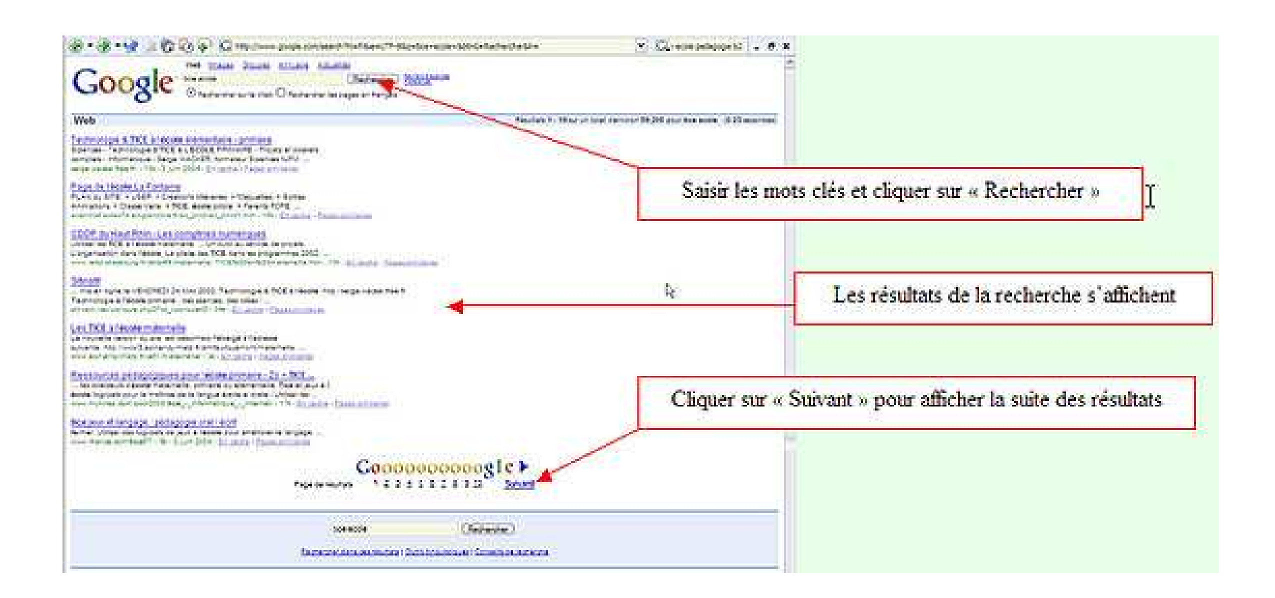
Nom	**Google**

URL	http://www.google.com/ http://www.google.fr/
Taille de l'index	plusieurs milliards de pages
Délai de rafraîchissement de l'index	4 semaines (période appelée la "Google Dance")
Syntaxe d'interrogation	
Minuscules / majuscules	**Idem** : Google ne tient pas compte de la casse des lettres. Exemple : ibm trouve ibm, IBM ou Ibm. Ibm trouve Ibm, ibm ou IBM. IBM trouve IBM, ibm ou Ibm.
Lettres accentuées	**Pas d'importance.** Exemple : electricite trouve electricite, électricité et electricité. électricité trouve électricité, electricite et electricité. Il est possible de rechercher une orthographe spécifique grâce au signe +. Exemple : +electricité**.**
Ordre des mots	**Important** : paris dakar donne un résultat différent de dakar paris. Une plus grande importance est donnée au premier mot choisi.
OU	**OR** Exemple : engine OR directory. L'opérateur doit être saisi en majuscules obligatoirement.
ET	**Opérateur par défaut** Exemple : moteur recherche recherche les pages qui contiennent à la fois moteur ET recherche. Il est également possible d'utiliser le signe + : Exemple : +moteur +recherche
SAUF	**-** Exemple : moteur -automobile recherche les pages qui contiennent moteur mais qui ne contiennent pas automobile.
Expressions	**Guillemets.** Exemple : "moteur de recherche" recherche cette expression, les trois mots les uns à côté des autres et dans cet ordre.
Synonymes	**Tilde (~)** Une recherche sur powerpoint ~help effectuera une recherche sur powerpoint ET help ou tips, faq, tutorial, etc.
Recherche sur le nom de domaine du site	**site:** Exemple : +moteur +site:www.honda.com recherchera les documents qui contiennent le mot moteur et qui sont disponibles sur le site www.honda.com.
Recherche de sites similaires	**related:** Exemple : related:www2.ac-lyon.fr/services/cdri/ trouvera des sites similaires au site du CDRI.
Recherche dans le titre (1)	**allintitle:** Exemple : allintitle:tice ecole trouvera les pages qui contiennent les mots tice ET ecole dans leur titre.
Recherche dans le titre (2)	**intitle:** Exemple : intitle:tice ecole trouvera les pages qui contiennent le mot tice dans leur titre et le mot ecole dans la page.
Recherche dans l'url (1)	**allinurl:** Exemple : allinurl:tice ecole trouvera les pages qui contiennent les mots tice ET ecole dans leur adresse (url).
Recherche dans l'url (2)	**inurl:** Exemple : inurl:ressources tice ecole trouvera les pages qui contiennent le mot ressources dans leur url et les mots tice et ecole dans la page.

Les résultats d'une recherche avec Google

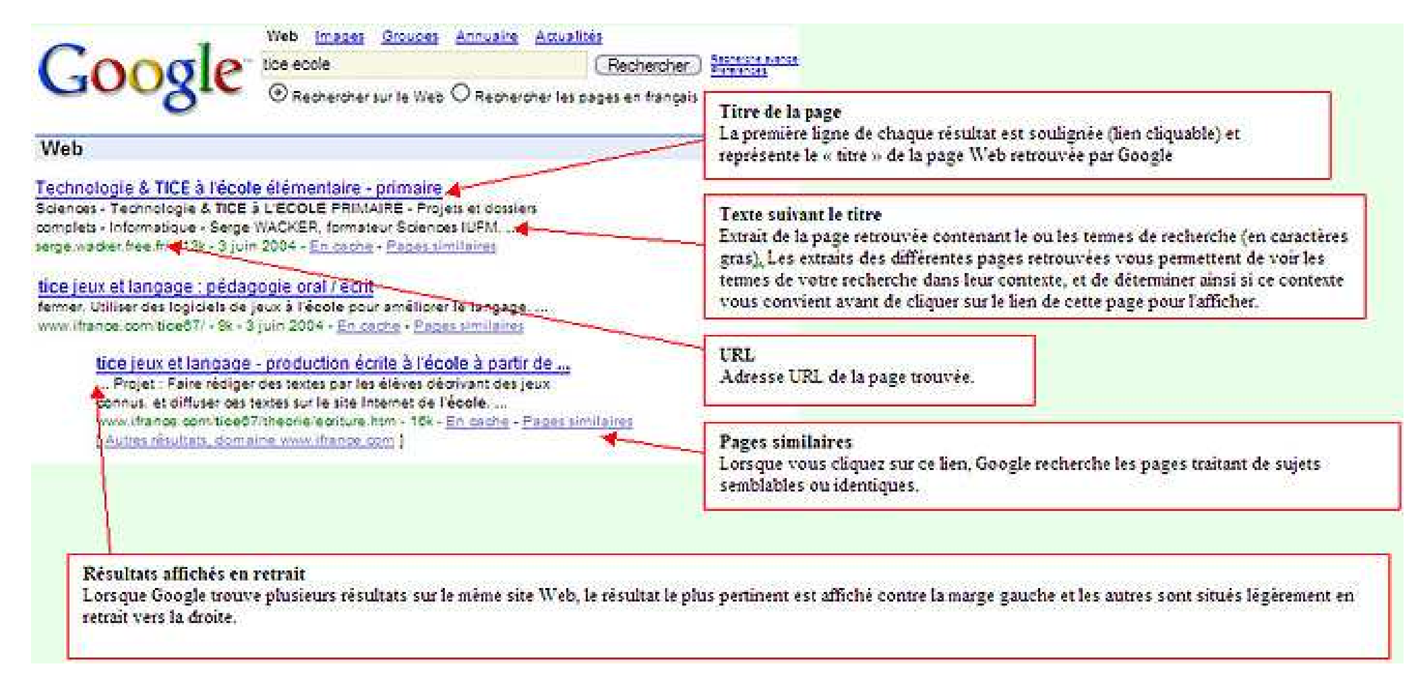
1. Rechercher et afficher les résultats

Dans votre navigateur, saisir l'adresse de Google (http://www.google.fr/), puis :



2. Analyse des résultats

La page de résultats comporte, pour chaque page trouvée, quelques éléments expliqués ci-dessous:
(le bouton « J'ai de la chance » affiche directement (et uniquement) la page Web considérée par Google comme la plus pertinente pour votre requête).



Un guide complet: http://www.lesmoteursderecherche.com/

Retour